Weighting components of teacher evaluation models

Laura Goe, Ph.D.

Research Scientist, ETS

Principal Investigator for Research and Dissemination, The National Comprehensive Center for Teacher Quality

Arizona Summit III

Arizona Department of Education

April 30, 2012 ♦ Phoenix, AZ

The goal of teacher evaluation

*The **ultimate** goal of all teacher evaluation should be…*

TO IMPROVE TEACHING AND LEARNING

To be discussed…

- Commonly used measures
- Weighting measures
- How the data is/will be used to inform decisions and *improve teaching and learning*

Values

- Including a measure in an evaluation model signals that you *value* the thing being measured
  - Observations: classroom practice is *valued*
  - Student growth measures (value-added, Colorado Growth model): student learning growth is *valued*
  - Student surveys: student perceptions are *valued*
  - Portfolios: teacher input and self-reflection are *valued*
  - Artifacts: lesson plans, assignments and student work are sources of *valued* information about teaching

Weights signal *value*

- Weights signal *how much* you *value* the thing being measured
  - Higher weights signal that you value that component more
    - Many systems are weighted more heavily on student outcomes than teacher practice
  - Lower weights signal that you value that component less
    - Surveys, portfolios and artifacts typically have lower weights in evaluation models

Weights signal *confidence*

- Weights signal how *confident* you are in the *accuracy* (validity) of the measure
  - Higher weights signal that you are more confident that the measure is accurately reflecting teacher effectiveness
    - "Objective" measures such as growth models based on standardized tests tend to be considered more accurate
    - "Subjective" measures such as observations and portfolios tend to be considered less accurate

How will teachers respond to weights?

- They will likely focus their attention on doing well on the measures that carry a higher weight in the evaluation model
- Examples of unintended consequences
  - Narrowing of the curriculum
  - Too much time spent on test prep
- Because of this concern, think carefully about the impact the weights will have on teacher practices

"In the fall, you will meet with your administrator to decide which assessment(s) you will use to evaluate your students' achievement. **If you are using multiple assessments, you will decide how to weight them.** Finally, you will also decide on your specific student learning targets for the year. Please note that **your administrator must approve your choice of assessments, the weights you assign to them, and your achievement targets.** Please also note that your administrator may choose to meet with groups of teachers from similar content areas rather than with each teacher individually."

Two approaches to combining measures

- AIR's Sheri Frost Leo and Lisa Lachlan-Haché (2012) have written a really useful paper on combining and weighting measures

- Two approaches are defined

  - Numerical approach: measures of teacher performance are quantified and added or averaged into a teacher effectiveness "score"

  - Profile approach: performance data are gathered and maintained separately, without adding or averaging the results across metrics; then placed into rating categories for each of the measures

An example of the numerical approach: Tennessee overall score calculator

Component	Score		Points
TEAM Score Calculator			TEAM TENNESSEE
Enter scores for each of the components of TEAM in the light blue cells to yield an Overall Effectiveness Rating			
Overall Observation Score*	3.3	x 50	165
Growth Score	3	x 35	105
15% Achievement Measure Score	4	x 15	60
		Total Score from the 3 Components =	330
Overall Effectiveness Rating	3		

*Overall observation scores for educators assessed according to the TEAM rubric are an average of 41 indicators scored for Professional teachers (3 Planning indicators, 4 Environment indicators, 24 Instruction indicators across two observations, and 10 Professionalism indicators) or an average of 60 indicators scored for Apprentice teachers (6 Planning indicators across 2 observations, 8 Environment indicators across two observations, 36 Instruction indicators across 3 observations, and 10 Professionalism indicators). These averages are rounded to the hundredth place. Use the "TEAM Observation Tracker" tab of this spreadsheet to calculate a running average. Growth scores and achievement measure scores are reported as whole numbers (1-5).

Total Scores are converted to an Overall Effectiveness Rating using this scale:	Total Score	Overall Effectiveness Rating
	Less than 200	1
	200–	2
	275–	3
	350–	4
	425-500	5

- Overall Observation Score x 50
- Growth Score x 35
- Achievement Measure Score x 15

- Overall Effectiveness Rating
  - 1 = Less than 200
  - 2 = 200+
  - 3 = 275+
  - 4 = 350+
  - 5 = 425-500

An example of the Profile approach: New Haven matrix

		Student Learning Growth				
		1	2	3	4	5
Instructional Practice and Professional Values	1	1	1	2	3*	3*
	2	1	2	2	3	4*
	3	1	2	3	4	5
	4	2*	3	4	4	5
	5	3*	3*	4	5	5

Asterisks indicate a mismatch—teacher is very high on one area (practice or growth) and very low on the other area.

Combining multiple student growth scores

- In evaluation models using student learning objectives, there may be multiple measures of student learning for an individual teacher
- An example for a 3rd grade teacher
    - A score for writing (using rubric)
    - A score for reading (using DIBELS)
    - A score for math & reading (using NWEA MAP)
    - An art portfolio (using rubric)

Final thoughts

- Remember that teachers will pay attention to what gets measured, so be sure your measures reflect the state's goals and values

- Remember that teachers will pay *more* attention to what gets weighted *more*, so consider the intended and unintended consequences of your weighting system

References & Resources

Colorado Content Collaboratives
http://www.cde.state.co.us/ContentCollaboratives/index.asp

Indiana RISE system www.RISEindiana.org

Leo, S. F., & Lachlan-Haché, L. (2012). Creating summative educator effectiveness scores: Approaches to combining measures. Washington, DC: American Institutes for Research.

Louisiana Student Growth for Non-tested Subjects
http://www.louisianaschools.net/compass/sgm_nontested.html

New York State Guidance on Student Learning Objectives

http://engageny.org/wp-content/uploads/2012/03/slo-guidance.pdf

Rhode Island Department of Education Teacher Evaluation – Student Learning Objectives
http://www.ride.ri.gov/educatorquality/educatorevaluation/SLO.aspx

Tennessee Teacher Evaluation http://team-tn.org/

Tripod Survey http://www.tripodproject.org/index.php/index/

Washington, DC IMPACT Evaluation system http://www.dc.gov/DCPS/impact

Laura Goe, Ph.D.
609-619-1648
lgoe@ets.org
www.lauragoe.com
https://twitter.com/GoeLaura

National Comprehensive Center for Teacher Quality
1000 Thomas Jefferson Street, NW
Washington, D.C. 20007
www.tqsource.org